An Automated Computerized Severity Index

Richard F. Gibson, M.D. and Peter J. Haug, M.D.
Department of Medical Informatics, LDS Hospital/University of Utah, Salt Lake City, Utah

The Computerized Severity Index (CSI) is a commercially available scoring system for hospital inpatients. Trained abstractors review the patient's paper medical record and enter the diagnoses and relevant physiological attributes. The HELP (Health Evaluation through Logical Processing) System at LDS Hospital stores patient data in discrete codes. This paper describes the development of an automatic interface between the standalone, personal-computer-based severity system and the mainframe-based hospital information system. The interface scores patient severity without the need for manual chart review. Severity scores from the automated and manual methods were identical for 70% of 222 general medical patients scored retrospectively. An evaluation of the causes for differing scores between the two methods is presented.

INTRODUCTION

Severity of illness indices are intriguing because they attempt to quantify the intangible feeling that providers have about the seriousness of their patients' illnesses. The long term goal of the authors is to completely automate one such index, the Computerized Severity Index (CSI), as an application in the HELP System using the LDS Hospital patient database. A prior abstract summarized the correlation between the data elements needed by CSI and the data elements defined in the HELP System [1]. A previous paper examined the proportion of 1356 patients with electronic data elements needed for CSI scoring [2].

We have developed an electronic interface between CSI and the HELP System which produces CSI scores without manual paper chart review. The dominant characteristic of the interface is the implementation of logic to correlate the atomic detail of patient data in the HELP System with the abstract concepts required by CSI.

SEVERITY INDICES

Severity of illness indices for hospitalized patients have been studied for more than 10 years. Such indices have been used to predict mortality [3-5], and explain variation in cost and length of stay among patients with similar diagnoses [5-7]. Severity of illness indices have also been used in quality assurance activities [8], and could be used to stratify patients entering clinical trials.

Two severity systems other than CSI use clinical data (as opposed to discharge abstract data) to calculate severity. Both systems use the same clinical data regardless of diagnoses. APACHE III (Acute Physiology and Chronic Health Evaluation) evaluates age, 7 comorbid conditions, and 17 physiologic variables recorded in the patient's chart. Although largely used to predict mortality for patients in the intensive care unit (ICU) [4], APACHE III has been used to explain variation in ICU length of stay [9], and APACHE II was used to explain variation in resource use [6].

MedisGroups (Medical Illness Severity Grouping System) assigns a score (0 to 3) to 250 key clinical findings (KCFs) abstracted from the medical chart. The overall admission score (0 to 4) is computed from the KCF scores. A mid-stay score is also derived using modified KCFs [10]. MedisGroups has been used to predict hospital charges [11] and mortality [12].

There are no literature reports of automatic data collection and score calculation for these two severity systems.

COMPUTERIZED SEVERITY INDEX (CSI)

CSI was developed by Susan Horn, Ph.D., and 200 nurses and physicians at The Johns Hopkins University and Hospital. CSI maps each of the approximately 12,000 ICD-9-CM diagnoses to one of 833 disease groups [13]. Each disease group is comprised of 4-50 indicators: physiological patient attributes such as vital signs, physical exam findings, and diagnostic studies. With rare exceptions, CSI uses no treatment or intervention facts in calculating severity. Severity scores from 1 to 4 are calculated for each indicator, for each disease group, and for the patient overall. A score of 1 indicates normal or mildly abnormal findings. A score of 4 indicates catastrophic or life-threatening signs or symptoms.

Closely related ICD-9-CM diagnoses, such as the various types of bacterial pneumonia, are mapped to the same disease group. A given patient may have 1 to 12 or more disease groups used, depending on the number and independence of the diagnoses. Different indicators are used for each

0195-4210/94/$5.00

Table 1: Some Pneumonia Indicators

Temperature	Lowest systolic blood pressure
Dyspnea	Highest white blood cell count
Cyanosis	Lowest oxygen level
Rales on lung exam	Chest X-ray findings

disease group, depending on the disease. A partial listing of the indicators for the pneumonia disease group is given in Table 1.

Each indicator is assigned a score from 1 to 4 based on the severity of the patient attribute. The score given to the same indicator may vary from disease group to disease group. For example, a temperature of 39°C may be a Score 3 in leukemia, but it is only a Score 2 in pneumonia. Table 2 shows two indicators for the pneumonia disease group.

The lower score of the two highest indicator scores determines the score of the disease group. Each indicator can be used only once per patient regardless of the number of the patient's disease groups using that indicator.

In manually scoring a hospitalization, a trained abstractor enters the diagnoses into the computer. The CSI program presents a list of indicators for all the disease groups suggested by the diagnoses. The abstractor then reviews the chart and notes the most extreme patient attributes relating to the indicators. The abstractor then chooses the level of each indicator. For descriptive indicators such as dyspnea (shortness of breath), the abstractor chooses the appropriate menu item. For numerical findings such as lowest systolic blood pressure, the abstractor enters the actual value of the attribute. After completion of the data entry, the computer calculates the severity score for each disease group, and for the patient overall. CSI is typically scored for Admission, Maximum, and Discharge. The Admission and Discharge periods are usually 24 hours after admission and before discharge, respectively. The Maximum score represents the contribution of the highest score of each indicator at any time during the hospitalization without regard to whether the indicators reached their highest scores concurrently. CSI can be tailored to calculate a score for any additional period, such as for admission or discharge from the ICU.

CSI is written in Advanced Revelation for

Table 2: Indicator Scoring for Pneumonia

	Dyspnea	Lowest Systolic BP
Score 1	no dyspnea	> 89
Score 2	dyspnea on exertion	80 - 89
Score 3	dyspnea at rest	61 - 79
Score 4	periods of apnea	< 61

standalone or networked IBM-compatible personal computers (PCs). Abstractors require 15-30 minutes to review the chart and enter the data, depending on the length and complexity of the hospitalization.

THE HELP SYSTEM

Elements of the HELP System have been under development at LDS Hospital, a 520-bed tertiary care center in Salt Lake City, since 1967 [14,15]. HELP provides an integrated, computerized environment for use and development of clinical, administrative, and financial modules. An integrated expert system tool is used to support medical decision making.

HELP uses a hierarchical, numerically-based coding scheme to represent medical terms. Drug names, laboratory tests, diagnoses, admission-discharge-transfer data, physical exam findings, and nursing care plans and actions are all represented by 8-byte codes called PTXT (pronounced "P-text", for Pointer-to-TeXT) defined in a comprehensive data dictionary. There are almost no PTXT codes for patient findings observed by physicians. Despite efforts to restrict new entries and discard unused PTXT codes, there are medical terms linked to more than one PTXT code, and medical events sometimes represented by a single PTXT code and sometimes by a cluster of PTXT codes.

The HELP System runs on a cluster of Tandem mainframes using a proprietary language. Patient data are compacted into "packed strings" which are stored on the permanent media in non-relational format by patient number and data class (the top level of the dictionary hierarchy). One thousand PCs connect with the Tandem via fiber and Ethernet.

THE HELP TO CSI INTERFACE

At present, the automated CSI system uses 4 steps which are manually started. (1) A C program runs on the mainframe and collects into an ASCII file all data that could possibly be used by CSI. After retrieving demographic data, admission and discharge times, and all ICD-9-CM diagnoses, the program collects multiple lines of a fixed format: a PTXT code, a data value (such as the actual blood pressure), and a timestamp. (2) The ASCII file is then transferred from the Tandem to a PC. (3) An interface program, written in Advanced Revelation, then sorts and analyzes the PTXT codes, and stores the most extreme indicator values in the CSI tables.

Table 3: Observed Scores of 222 Test Cases

Auto	Manual CSI Score				
Score	1	2	3	4	
1	74	21	3	1	99
2	7	30	8	5	50
3	0	6	17	10	33
4	0	2	3	35	40
	81	59	31	51	222

(4) The user then enters the regular CSI application program which calculates the severity score.

Nurses enter their observations into the HELP System patient database by choosing menu items and then applying a timestamp. They record the patient findings in an atomic fashion, one finding at a time, such as "chest pain at 1530h". HELP stores this event as 3 PTXT codes: pain, location (anterior chest), pain intensity (on a scale of 1 to 10), and a timestamp. CSI calculates severity scores based on more abstract concepts such as "recurrent severe chest pain". The interface program encodes the logic to map between the atomic PTXT codes and the abstract CSI concepts. In the above example, CSI looks for a pain code, an anterior chest location code, an intensity code with a data value of 7-10, and then checks all three codes for the same timestamp. If CSI can find 4 or more of these code clusters in a 24-hour period then the indicator is considered satisfied at a score of 4.

Medical judgment was necessary to equate the different vocabularies of the two database systems. There are no PTXT codes to represent many of the CSI indicators, but PTXT codes do exist for most of the common indicators (such as temperature and white blood cell count) used in many of the disease groups [1,2].

METHODS

The manual version of CSI has been in use for 2 years at LDS Hospital. In 1993-4 CSI coders manually scored the charts of 2000 patients with pressure sores for use in another clinical study. Of those 2000 charts, a convenience set of 352 was used to develop and test the automatic interface: 130 cases in the training set and 222 cases in the test set. (The most recently scored 352 cases were chosen because the CSI scoring engine has been modified since 1993 and we wanted to avoid scoring discrepancies due to different versions of the CSI software.) For the 130 patients in the training set, an iterative process involved automatically scoring 5 patients, examining the differences in the automated and manual scores, modifying the interface, and then scoring 5 more cases. The instrument was then frozen and used to obtain automated CSI scores on the test set.

In this formative study, no attempt was made to obtain a representative sample of LDS Hospital discharges or to control for diagnosis, hospital division, or length of stay. However, the patients in the pressure sore study exhibited a wide range of medical and surgical diagnoses and all hospital divisions with HELP nurse charting were represented. There is no HELP nurse charting on the rehabilitation, psychiatric, or obstetric wards, and these patients were excluded from study.

The 2 independent variables in this study are the automated and manual CSI scoring methods. The dependent variable is the CSI score. Interrater agreement was measured using the Kappa statistic [16], weighted Kappa [16], intraclass correlation [17], and Finn's intraclass correlation [17]. For weighted Kappa weights we used the square of the difference between CSI scores $(\text{automated-manual})^2$.

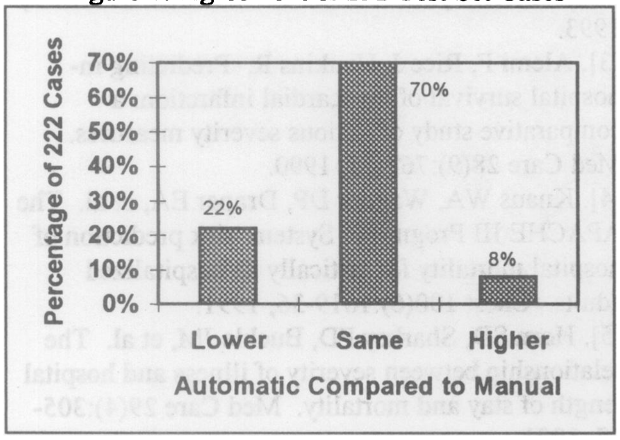
RESULTS

Table 3 shows the scores of the 222 test set patients. The observed agreement was 70.3% (156 of 222). The Kappa statistic was $\kappa=.584$ ($p<.001$). Weighted Kappa was $\kappa_W=.823$. The intraclass correlation coefficient was $R_I=.951$ ($p<.001$). Finn's adjusted intraclass correlation was $r_F=.813$.

Figure 1 shows how the automated scores compared to the manual CSI scores. Figure 2 displays the distribution of cases grouped by manual CSI score. Figure 3 shows agreement classified by manual CSI score.

Only the training set was examined case-by-case to determine the reason for differing scores. In the 16 cases (12% of the 130 training cases) where the automatic score was higher than the manual score, there were 8 instances of manual coder error (mostly overlooking a data element). There were 7 instances when a patient finding fell into a post-operative "window". For example, CSI coders ignore hematuria for 48 hours after bladder surgery, but this logic has not yet been encoded into the automated CSI. There were 2 cases where the manual coder exercised judgment in ignoring what appeared to be artifactually low blood pressures. Such logic has not yet been encoded into the electronic interface. There were 2 cases where portions of the laboratory studies and vital signs had not been printed for the summary-to-date report in the paper chart so the manual coders did not find the

Figure 1: Agreement of 222 Test Set Cases



abnormal indicators. Some patients exhibited more than one reason for differing scores. If the 8 cases of coder error are removed from the "high" category, the observed agreement of automated and manual scores rises from 73% to 79% (103 of 130 cases).

Examination of the 19 cases (15% of 130 training cases) where the automated score was lower than the manual score revealed 13 instances where PTXT codes do not exist for the indicator (for example, the number of lower extremity fractures). There were 9 cases where existing PTXT codes were not used by the nursing staff. There were 2 cases with a manual coder error. Only one instance was found where the indicator was derived from physician dictation ("swollen prostate") and thereby missed by automated CSI (no PTXT code). Some cases had more than one reason for the difference in scores. Table 4 highlights the causes of disagreement between automated and manual CSI.

Although the test set cases with unlike scores were not examined, it is likely that the causes for differing automated and manual scores were the same for the test and training sets. The distribution of cases by CSI score was similar, as was the observed agreement between automated and manual

Figure 2: Distribution of 222 Test Set Cases

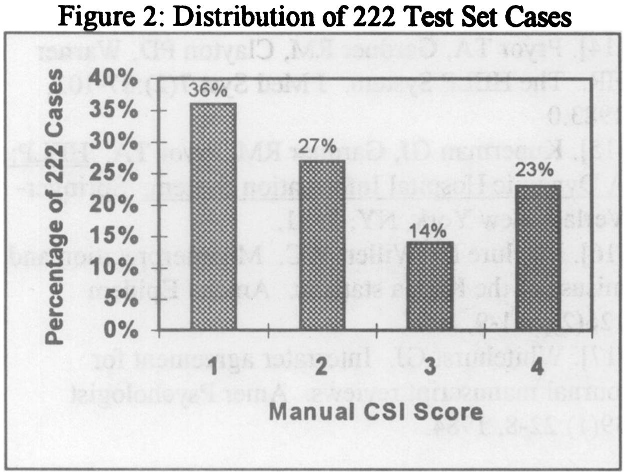
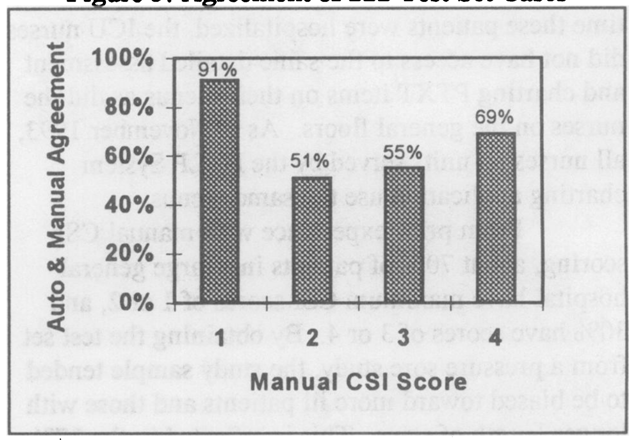


Figure 3: Agreement of 222 Test Set Cases



scores (73% in the training set, 70% in the test set).

DISCUSSION

Automated CSI scoring resulted in scores both higher and lower than manual scoring. An advantage of an electronic interface is the avoidance of data collection errors inherent in manual review of complex cases, responsible for half of the higher automated scores in the training set. Most of the remaining higher automated scores were due to patient findings in the post-operative window. Our next task is to modify the automated CSI interface to ignore patient findings in the post-operative window.

After successful encoding of the post-operative window, there is promise that the automated CSI score could become a *minimum* CSI score. That is, the manual score would never be lower than the automated score. If necessary, manual coders could then perform a brief, directed search for indicators known to have no corresponding PTXT codes.

This study demonstrates that manual coders found some patient attributes not yet computerized in the HELP System. Most of these attributes could be collected by nurses if the HELP System menus were modified. Other attributes were computerized, but,

Table 4: Causes of Auto vs. Manual Disagreement

Causes for auto CSI to be higher than manual:

- automated CSI had no post-op window
- manual coder error
- manual coder rejected artifactual data
- paper chart incomplete

Causes for auto CSI to be lower than manual:

- PTXT codes do not exist for patient finding
- nurses did not use existing PTXT code
- manual coder error

in some cases, were not used by the nurses. At the time these patients were hospitalized, the ICU nurses did not have access to the same detailed assessment and charting PTXT items on their menus as did the nurses on the general floors. As of November 1993, all nurses on units served by the HELP System charting application use the same menus.

From prior experience with manual CSI scoring, about 70% of patients in a large general hospital have maximum CSI scores of 1 or 2, and 30% have scores of 3 or 4. By obtaining the test set from a pressure sore study, the study sample tended to be biased toward more ill patients and those with longer length of stays. This is reflected in the 37% of test set patients with manual CSI scores of 3 or 4.

Further study of automated CSI scoring will involve larger sample sizes. Correlation of automated and manual scores will be controlled for hospital division, length of stay, and the Major Diagnostic Group of the principal diagnosis. Examination of more cases where the automated scores are low could suggest additions to the HELP System computerized nurse charting.

CONCLUSION

It appears that a detailed clinical patient database such as the HELP System can drive an automated severity of illness index in areas of the hospital served by the nurse charting system. Significant challenges remain in encoding logic used by manual abstractors to ignore inappropriate data values and expected abnormal patient findings. As the HELP System and other hospital information systems add coded clinical data, automated CSI scoring can be expected to improve.

ACKNOWLEDGMENTS

This publication was supported by grant number F37 LM00012 from the National Library of Medicine. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Library of Medicine. The authors are indebted to Blair C. James for his programming assistance.

Reference

[1]. Gibson RF, Haug PJ, Horn SD, Gardner RM. Linking Computerized Severity Index (CSI) with the HELP System patient database. Abstract, Spring Congress, AMIA, 1993.

[2]. Gibson R, Haug P. Linking the Computerized Severity Index (CSI) to coded patient findings in the HELP System patient database. SCAMC 17:673-7, 1993.

[3]. Alemi F, Rice J, Hankins R. Predicting in-hospital survival of myocardial infarction: a comparative study of various severity measures. Med Care 28(9):762-75, 1990.

[4]. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III Prognostic System: risk prediction of hospital mortality for critically ill hospitalized adults. Chest 100(6):1619-36, 1991.

[5]. Horn SD, Sharkey PD, Buckle JM, et al. The relationship between severity of illness and hospital length of stay and mortality. Med Care 29(4):305-17, 1991.

[6]. Thomas JW, Ashcraft MLF. Measuring severity of illness: six severity systems and their ability to explain cost variations. Inquiry 28(1):39-55, 1991.

[7]. Averill RF, McGuire TE, Manning BE, et al. A study of the relationship between severity of illness and hospital cost in New Jersey hospitals. Health Services Research 27(5):587-604, 1992.

[8]. Iezzoni LI. Using severity information for quality assessment: a review of three cases by five severity measures. QRB 15(12):376-82, 1989.

[9]. Knaus WA, Wagner DP, Zimmerman JE, Draper EA. Variations in mortality and length of stay in intensive care units. Ann Int Med 118(10):753-61, 1993.

[10]. Iezzoni LI. A primer on MedisGroups. PA Med 92(1):28-33, 1989.

[11]. Iezzoni LI, Ash AS, Coffman GA, Moskowitz MA. Admission and mid-stay MedisGroup scores as predictors of hospitalization charges. Med Care 29(3):210-20, 1991.

[12]. Iezzoni LI, Ash AS, Coffman G, Moskowitz MA. Admission and mid-stay MedisGroups scores as predictors of death within 30 days of hospital admission. Am J Pub Health 81(1):74-8, 1991.

[13]. Iezzoni LI, Daley J. A description and clinical assessment of the Computerized Severity Index. Qual Rev Bull 18(2):44-52, 1992.

[14]. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP System. J Med Syst 7(2):87-102, 1983.0

[15]. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. Springer-Verlag, New York, NY, 1991.

[16]. Maclure M, Willett WC. Misinterpretation and misuse of the Kappa statistic. Amer J Epidem 126(2):161-9, 1987.

[17]. Whitehurst GJ. Interrater agreement for journal manuscript reviews. Amer Psychologist 39(1):22-8, 1984.